Pairwise Constraint Propagation on Multi-View Data

Zhiwu Lu and Liwei Wang

arXiv:1501.04284v1 [cs.CV] 18 Jan 2015

Abstract—This paper presents a graph-based learning approach to pairwise constraint propagation on multi-view data. Although pairwise constraint propagation has been studied extensively, pairwise constraints are usually defined over pairs of data points from a single view, i.e., only intra-view constraint propagation is considered for multi-view tasks. In fact, very little attention has been paid to inter-view constraint propagation, which is more challenging since pairwise constraints are now defined over pairs of data points from different views. In this paper, we propose to decompose the challenging inter-view constraint propagation problem into semi-supervised learning subproblems so that they can be efficiently solved based on graph-based label propagation. To the best of our knowledge, this is the first attempt to give an efficient solution to inter-view constraint propagation from a semi-supervised learning viewpoint. Moreover, since graph-based label propagation has been adopted for basic optimization, we develop two constrained graph construction methods for inter-view constraint propagation, which only differ in how the intra-view pairwise constraints are exploited. The experimental results in cross-view retrieval have shown the promising performance of our inter-view constraint propagation.

Index Terms—Pairwise constraint propagation, multi-view data, label propagation, graph construction, cross-view retrieval

I. INTRODUCTION

As an alternative type of supervisory information easier to access than the class labels of data points, pairwise constraints are widely used for different machine learning tasks in the literature. To effectively exploit pairwise constraints for clustering or classification [1]–[4], much attention has been paid to pairwise constraint propagation [5]–[7]. Different from the method [8] which only adjusts the similarities between constrained data points, these approaches can propagate pairwise constraints to other similarities between unconstrained data points and thus achieve better results in most cases. More importantly, given that each pairwise constraint is actually defined over a pair of data points from a single view, these approaches can all be regarded as intra-view constraint propagation when multi-view data is concerned. Since we have to learn the relationships (must-link or cannot-link) between data points, intra-view constraint propagation is more challenging than the traditional label propagation [9]–[14] whose goal is only to predict the labels of unlabeled data points.

However, besides intra-view pairwise constraints, we may also have easy access to inter-view pairwise constraints in multi-view tasks such as cross-view retrieval [15], where each pairwise constraint is defined over a pair of data points

Z. Lu is with the Key Laboratory of Data Engineering and Knowledge Engineering (MOE), School of Information, Renmin University of China, Beijing 100872, China (e-mail: zhiwu.lu@gmail.com).

L. Wang is with the Key Laboratory of Machine Perception (MOE), School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China (e-mail: wanglw@cis.pku.edu.cn).

from different views (see Fig. 1). In this case, inter-view pairwise constraints still specify the must-link or cannot-link relationships between data points. Since the similarity of two data points from different views is commonly unknown in practice, inter-view constraint propagation is significantly more challenging than intra-view constraint propagation. In fact, very little attention has been paid to inter-view constraint propagation for multi-view tasks in the literature. Although pairwise constraint propagation has been successfully applied to multi-view clustering in [16], [17], only intra-view pairwise constraints are propagated across different views. Here, it should be noted that these two constraint propagation methods *have actually ignored* the concept of inter-view pairwise constraints or the strategy of inter-view constraint propagation.

Since multi-view data can be readily decomposed into a series of two-view data, we focus on inter-view constraint propagation only across two views in this paper. However, such inter-view constraint propagation remains a rather challenging task. Fortunately, from a semi-supervised learning viewpoint, we can formulate inter-view constraint propagation as minimizing a regularized energy functional. Specifically, we first decompose the inter-view constraint propagation problem into a set of independent semi-supervised learning [9]–[12] subproblems. Through formulating these subproblems uniformly as minimizing a regularized energy functional, we thus develop an efficient algorithm for inter-view constraint propagation based on the traditional graph-based label propagation technique [9]. In summary, we succeed in giving an insightful explanation of inter-view constraint propagation from a graph-based semi-supervised learning viewpoint.

However, since graph-based label propagation has been adopted for basic optimization, there remains one problem to be concerned in inter-view constraint propagation, i.e., how to exploit intra-view pairwise constraints for graph construction within each view. In this paper, we develop two constrained graph construction methods for inter-view constraint propagation, which only differ in how the intra-view pairwise constraints are exploited. The first method limits our inter-view constraint propagation to a single view and then utilize the constraint propagation results to adjust the weight matrix of each view, while the second method formulates graph construction as sparse representation and then directly add the intra-view pairwise constraints into sparse representation.

The flowchart of our inter-view constraint propagation with constrained graph construction is illustrated in Fig. 1, where only two views (i.e. text and image) are considered. It should be noted that, when multiple views refer to text, image, audio and so on, the output of our inter-view constraint propagation actually denotes the correlation between different media views. That is, the proposed algorithm can be directly used for cross-view retrieval (also see examples in Fig. 3) which has
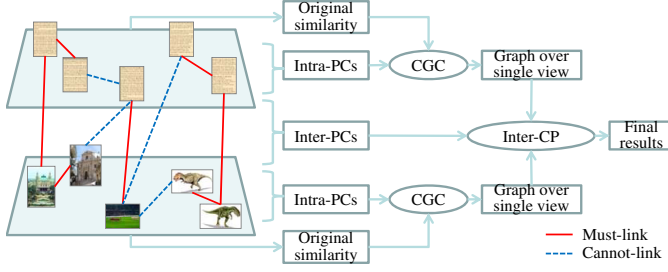
Fig. 1. Illustration of the flowchart of inter-view constraint propagation (Inter-CP) with constrained graph construction (CGC). Here, we only consider two different views: text and image. Moreover, Intra-PCs and Inter-PCs denote intra-view and inter-view pairwise constraints, respectively.

drawn much attention recently [15]. For cross-view retrieval, it is not feasible to combine multiple views just as previous multi-view retrieval methods [18], [19]. More notably, the two closely related methods [16], [17] for multi-view clustering are actually incompetent for cross-view retrieval.

Finally, to emphasize our main contributions, we summarize the following distinct advantages of our pairwise constraint propagation on multi-view data:

- We have made the first attempt to give an efficient solution to inter-view constraint propagation from a graph-based semi-supervised learning viewpoint.
- We have developed two constrained graph construction methods so that the intra-view pairwise constraints can also be exploited for inter-view constraint propagation.
- When applied to cross-view retrieval, our inter-view constraint propagation has been shown to achieve promising results with respect to the state-of-the-art.
- Although only evaluated in cross-view retrieval, our inter-view constraint propagation can be readily extended to many other multi-view tasks.

The remainder of this paper is organized as follows. In Section II, we formulate inter-view constraint propagation from a semi-supervised learning viewpoint. In Section III, we develop two constrained graph construction methods for our inter-view constraint propagation. In section IV, our inter-view constraint propagation is applied to cross-view retrieval. Finally, Sections V and VI provide the experimental results and conclusions, respectively.

II. INTER-VIEW CONSTRAINT PROPAGATION

In this section, we first formulate inter-view constraint propagation as minimizing a regularized energy functional from a semi-supervised learning viewpoint. Furthermore, we develop an efficient algorithm for inter-view constraint propagation based on the label propagation technique [9].

A. Problem Formulation

Given a set of inter-view pairwise constraints defined over pairs of data points from different views, the goal of inter-view constraint propagation is to learn the cross-view relationships from these initial pairwise constraints. Since the similarity of two data points from different views is unknown in practice, inter-view constraint propagation on multi-view data is much

more challenging than the traditional pairwise constraint propagation over a single view. Considering that this multi-view problem can be readily decomposed into a series of two-view subproblems, we focus on inter-view constraint propagation on two-view data in the following.

Let $\{\mathcal{X}, \mathcal{Y}\}$ be a two-view dataset, where $\mathcal{X} = \{x_1, ..., x_N\}$ and $\mathcal{Y} = \{y_1, ..., y_M\}$. It should be noted that we may have $N \neq M$. As an example, a two-view dataset is shown in Fig. 1, with image and text being the two different views. For the two-view dataset $\{\mathcal{X}, \mathcal{Y}\}$, we can define a set of initial must-link constraints as $\mathcal{M} = \{(x_i, y_j) : l(x_i) = l(y_j)\}$ and a set of initial cannot-link constraints as $\mathcal{C} = \{(x_i, y_j) : l(x_i) \neq l(y_j)\}$, where $l(x_i)$ (or $l(y_j)$) is the class label of $x_i \in \mathcal{X}$ (or $y_j \in \mathcal{Y}$). Here, the two data points x_i and y_j are assumed to share the same class label set. If the class labels are not provided, the inter-view pairwise constraints can be defined only based on the correspondence between two views, which can be readily obtained from Web-based content (e.g. Wikipedia articles). Several examples of inter-view pairwise constraints are illustrated in Fig. 1.

We can now state that the goal of inter-view constraint propagation is to propagate the two sets of initial pairwise constraints $\mathcal{M}$ and $\mathcal{C}$ across both $\mathcal{X}$ and $\mathcal{Y}$. In fact, this is equivalent to deriving the best solution $F^* \in \mathcal{F}$ from both $\mathcal{M}$ and $\mathcal{C}$, with $\mathcal{F} = \{F = \{f_{ij}\}_{N \times M}\}$. Here, any exhaustive set of inter-view pairwise constraints is denoted as $F \in \mathcal{F}$, where $f_{ij} > 0$ means (x_i, y_j) is a must-link constraint while $f_{ij} < 0$ means (x_i, y_j) is a cannot-link constraint, with $|f_{ij}|$ denoting the confidence score of (x_i, y_j) being a must-link (or cannot-link) constraint. Hence, $\mathcal{F}$ can actually be regarded as the feasible solution set of inter-view constraint propagation.

Although it is difficult to directly find the best solution $F^* \in \mathcal{F}$ to inter-view constraint propagation, we can tackle this challenging problem by decomposing it into a set of independent semi-supervised learning subproblems. More concretely, we first denote the two sets of initial pairwise constraints $\mathcal{M}$ and $\mathcal{C}$ with a single matrix $Z = \{z_{ij}\}_{N \times M}$:

$$z_{ij} = \begin{cases} +1, & (x_i, y_j) \in \mathcal{M}; \\ -1, & (x_i, y_j) \in \mathcal{C}; \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Moreover, by making vertical and horizontal observations on such initial matrix Z, we decompose the inter-view constraint propagation problem into independent semi-supervised learning subproblems, which is also illustrated in Fig. 2. Finally, given two graphs $\mathcal{G}_{\mathcal{X}} = \{\mathcal{X}, W_{\mathcal{X}}\}$ and $\mathcal{G}_{\mathcal{Y}} = \{\mathcal{Y}, W_{\mathcal{Y}}\}$ constructed over $\{\mathcal{X}, \mathcal{Y}\}$ with $W_{\mathcal{X}}$ (or $W_{\mathcal{Y}}$) being the edge weight matrix defined over the vertex set $\mathcal{X}$ (or $\mathcal{Y}$), we utilize the graph-based label propagation method [9] to uniformly solve these semi-supervised learning subproblems:

$$\min_{F_{\mathcal{X}}, F_{\mathcal{Y}}} \|F_{\mathcal{X}} - Z\|_{fro}^2 + \mu_{\mathcal{X}} \mathrm{tr}(F_{\mathcal{X}}^T \mathcal{L}_{\mathcal{X}} F_{\mathcal{X}}) + \|F_{\mathcal{Y}} - Z\|_{fro}^2$$
$$+ \mu_{\mathcal{Y}} \mathrm{tr}(F_{\mathcal{Y}} \mathcal{L}_{\mathcal{Y}} F_{\mathcal{Y}}^T) + \gamma \|F_{\mathcal{X}} - F_{\mathcal{Y}}\|_{fro}^2, \quad (2)$$

where $\mu_{\mathcal{X}} > 0$ ($\mu_{\mathcal{Y}} > 0$, or $\gamma > 0$) denotes the regularization parameter, $\mathcal{L}_{\mathcal{X}}$ (or $\mathcal{L}_{\mathcal{Y}}$) denotes the normalized Laplacian matrix defined over $\mathcal{X}$ (or $\mathcal{Y}$), $\|\cdot\|_{fro}$ denotes the Frobenius norm of a matrix, and $\mathrm{tr}(\cdot)$ denotes the trace of a matrix.

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
x_1	0	-1	0	1	0	0	1	0
x_2	-1	0	1	-1	0	0	0	0
x_3	1	-1	0	0	0	-1	0	1
x_4	0	1	0	0	0	0	-1	0
x_5	0	0	0	1	0	0	0	-1

must-link: 1   cannot-link: -1

Fig. 2. Illustration of the initial matrix Z. When we focus on a single pair of data points, e.g. (x_3, y_4) here, the inter-view constraint propagation can be viewed as a two-class semi-supervised learning problem (in name only) in both vertical and horizontal directions, where +1 (or -1) denotes positive (or negative) labeled data and 0 denotes unlabeled data.

The first and second terms of the above objective function are related to the pairwise constraint propagation over $\mathcal{X}$, while the third and fourth terms are related to the pairwise constraint propagation over $\mathcal{Y}$. Moreover, the fifth term can ensure that the solutions of these two types of pairwise constraint propagation are as approximate as possible. Let $F_{\mathcal{X}}^*$ and $F_{\mathcal{Y}}^*$ be the best solutions of pairwise constraint propagation over $\mathcal{X}$ and $\mathcal{Y}$, respectively. The best solution of our inter-view constraint propagation is defined as follows:

$$F^* = (F_{\mathcal{X}}^* + F_{\mathcal{Y}}^*)/2. \tag{3}$$

As for the second and fourth terms, they are known as the energy functional [10] (or smoothness) defined over $\mathcal{X}$ and $\mathcal{Y}$. In summary, we have formulated intere-view constraint propagation as minimizing a regularized energy functional.

B. Efficient Algorithm

Let $\mathcal{Q}(F_{\mathcal{X}}, F_{\mathcal{Y}})$ denote the objective function in equation (2). The alternate optimization technique can be adopted to solve $\min_{F_{\mathcal{X}}, F_{\mathcal{Y}}} \mathcal{Q}(F_{\mathcal{X}}, F_{\mathcal{Y}})$ as follows: 1) Fix $F_{\mathcal{Y}} = F_{\mathcal{Y}}^*$, and find $F_{\mathcal{X}}^* = \arg\min_{F_{\mathcal{X}}} \mathcal{Q}(F_{\mathcal{X}}, F_{\mathcal{Y}}^*)$; 2) Fix $F_{\mathcal{X}} = F_{\mathcal{X}}^*$, and find $F_{\mathcal{Y}}^* = \arg\min_{F_{\mathcal{Y}}} \mathcal{Q}(F_{\mathcal{X}}^*, F_{\mathcal{Y}})$.

Pairwise Constraint Propagation over $\mathcal{X}$: When $F_{\mathcal{Y}}$ is fixed at $F_{\mathcal{Y}}^*$, the solution of $\min_{F_{\mathcal{X}}} \mathcal{Q}(F_{\mathcal{X}}, F_{\mathcal{Y}}^*)$ can be found by solving the following linear equation

$$\frac{\partial \mathcal{Q}(F_{\mathcal{X}}, F_{\mathcal{Y}}^*)}{2\partial F_{\mathcal{X}}} = (F_{\mathcal{X}} - Z) + \mu_{\mathcal{X}}\mathcal{L}_{\mathcal{X}}F_{\mathcal{X}} + \gamma(F_{\mathcal{X}} - F_{\mathcal{Y}}^*) = 0,$$

which can be equivalently transformed into:

$$(I + \hat{\mu}_{\mathcal{X}}\mathcal{L}_{\mathcal{X}})F_{\mathcal{X}} = (1 - \beta)Z + \beta F_{\mathcal{Y}}^*, \tag{4}$$

where $\hat{\mu}_{\mathcal{X}} = \mu_{\mathcal{X}}/(1+\gamma)$ and $\beta = \gamma/(1+\gamma)$. Since $I + \hat{\mu}_{\mathcal{X}}\mathcal{L}_{\mathcal{X}}$ is positive definite, we then obtain an analytical solution:

$$F_{\mathcal{X}}^* = (I + \hat{\mu}_{\mathcal{X}}\mathcal{L}_{\mathcal{X}})^{-1}((1 - \beta)Z + \beta F_{\mathcal{Y}}^*). \tag{5}$$

However, this analytical solution is not efficient for large datasets, since matrix inverse has a time cost of $O(N^3)$. Fortunately, equation (4) can also be *efficiently found using label propagation* [9] with k-nearest neighbor (k-NN) graph. **Pairwise Constraint Propagation over $\mathcal{Y}$:** When $F_{\mathcal{X}}$ is fixed at $F_{\mathcal{X}}^*$, the solution of $\min_{F_{\mathcal{Y}}} \mathcal{Q}(F_{\mathcal{X}}^*, F_{\mathcal{Y}})$ can be found by solving the following linear equation

$$\frac{\partial \mathcal{Q}(F_{\mathcal{X}}^*, F_{\mathcal{Y}})}{2\partial F_{\mathcal{Y}}} = (F_{\mathcal{Y}} - Z) + \mu_{\mathcal{Y}}F_{\mathcal{Y}}\mathcal{L}_{\mathcal{Y}} + \gamma(F_{\mathcal{Y}} - F_{\mathcal{X}}^*) = 0,$$

which can be equivalently transformed into:

$$F_{\mathcal{Y}}(I + \hat{\mu}_{\mathcal{Y}}\mathcal{L}_{\mathcal{Y}}) = (1 - \beta)Z + \beta F_{\mathcal{X}}^*, \tag{6}$$

where $\hat{\mu}_{\mathcal{Y}} = \mu_{\mathcal{Y}}/(1+\gamma)$ and $\beta = \gamma/(1+\gamma)$. Since $I + \hat{\mu}_{\mathcal{Y}}\mathcal{L}_{\mathcal{Y}}$ is positive definite, we then obtain an analytical solution:

$$F_{\mathcal{Y}}^* = ((1 - \beta)Z + \beta F_{\mathcal{X}}^*)(I + \hat{\mu}_{\mathcal{Y}}\mathcal{L}_{\mathcal{Y}})^{-1}, \tag{7}$$

which involves time-consuming matrix inverse. In fact, the linear equation (6) can also be efficiently solved using label propagation [9] with k-NN graph.

Let $W_{\mathcal{X}}$ (or $W_{\mathcal{Y}}$) denote the weight matrix of the k-NN graph constructed over $\mathcal{X}$ (or $\mathcal{Y}$). The complete algorithm for inter-view constraint propagation is summarized as follows:

(1) Compute two matrices $S_{\mathcal{X}} = D_{\mathcal{X}}^{-1/2}W_{\mathcal{X}}D_{\mathcal{X}}^{-1/2}$ and $S_{\mathcal{Y}} = D_{\mathcal{Y}}^{-1/2}W_{\mathcal{Y}}D_{\mathcal{Y}}^{-1/2}$, where $D_{\mathcal{X}}$ (or $D_{\mathcal{Y}}$) is a diagonal matrix with its i-th diagonal entry being the sum of the i-th row of $W_{\mathcal{X}}$ (or $W_{\mathcal{Y}}$);

(2) Initialize $F_{\mathcal{X}}(0) = 0$, $F_{\mathcal{Y}}^* = 0$, and $F_{\mathcal{Y}}(0) = 0$;

(3) Iterate $F_{\mathcal{X}}(t+1) = \alpha_{\mathcal{X}}S_{\mathcal{X}}F_{\mathcal{X}}(t) + (1-\alpha_{\mathcal{X}})((1-\beta)Z + \beta F_{\mathcal{Y}}^*)$ until convergence at $F_{\mathcal{X}}^*$, where $\alpha_{\mathcal{X}} = \hat{\mu}_{\mathcal{X}}/(1+\hat{\mu}_{\mathcal{X}})$ and $\beta = \gamma/(1+\gamma)$;

(4) Iterate $F_{\mathcal{Y}}(t+1) = \alpha_{\mathcal{Y}}F_{\mathcal{Y}}(t)S_{\mathcal{Y}} + (1-\alpha_{\mathcal{Y}})((1-\beta)Z + \beta F_{\mathcal{X}}^*)$ until convergence at $F_{\mathcal{Y}}^*$, where $\alpha_{\mathcal{Y}} = \hat{\mu}_{\mathcal{Y}}/(1+\hat{\mu}_{\mathcal{Y}})$;

(5) Iterate Steps (3)–(4) until convergence, and output the final solution $F^* = (F_{\mathcal{X}}^* + F_{\mathcal{Y}}^*)/2$.

According to the convergence analysis in [9], Step (3) converges to $F_{\mathcal{X}}^* = (1-\alpha)(I - \alpha_{\mathcal{X}}S_{\mathcal{X}})^{-1}((1-\beta)Z + \beta F_{\mathcal{Y}}^*)$, equal to the solution (5) given that $\alpha_{\mathcal{X}} = \hat{\mu}_{\mathcal{X}}/(1+\hat{\mu}_{\mathcal{X}})$ and $S_{\mathcal{X}} = I - \mathcal{L}_{\mathcal{X}}$. Similarly, Step (4) converges to $F_{\mathcal{Y}}^* = (1-\alpha)((1-\beta)Z + \beta F_{\mathcal{X}}^*)(I - \alpha_{\mathcal{Y}}S_{\mathcal{Y}})^{-1}$, equal to the solution (7) given that $\alpha_{\mathcal{Y}} = \hat{\mu}_{\mathcal{Y}}/(1+\hat{\mu}_{\mathcal{Y}})$ and $S_{\mathcal{Y}} = I - \mathcal{L}_{\mathcal{Y}}$. In the experiments, we find that Steps (3)–(5) generally converge in very limited iterations (<10). Moreover, based on k-NN graphs, the above inter-view constraint propagation algorithm has a time cost of $O(kNM)$, which is proportional to the number of all possible inter-view pairwise constraints. Hence, we consider that this algorithm can provide an efficient solution to inter-view constraint propagation (note that even a simple assignment operator on F^* incurs a time cost of $O(NM)$).

III. Constrained Graph Construction

In the last section, we have just developed an efficient inter-view constraint propagation algorithm based on the graph-based label propagation technique. However, since graph-based label propagation has been adopted as a basic optimization technique, there remains one problem to be concerned in inter-view constraint propagation, i.e., how to exploit intra-view pairwise constraints for graph construction within each view. In this section, we then develop two constrained graph construction methods for inter-view constraint propagation, which only differ in how the intra-view pairwise constraints are exploited. To ensure our inter-view constraint propagation algorithm runs efficiently even on large datasets, we utilize the traditional k-NN graph construction as the basis of our constrained graph construction, i.e., the obtained two constrained graphs can be considered as the variants of k-NN graph. In the

4

following, we will only elaborate how to construct the graph $\mathcal{G}_\mathcal{X} = \{\mathcal{X}, W_\mathcal{X}\}$ over $\mathcal{X}$. The graph $\mathcal{G}_\mathcal{Y} = \{\mathcal{Y}, W_\mathcal{Y}\}$ over $\mathcal{Y}$ can be constructed exactly in the same way.

A. Constrained Weight Adjustment

The first constrained graph construction method limits our inter-view constraint propagation proposed in Section II to a single view (i.e. intra-view constraint propagation over $\mathcal{X}$) and then utilize the obtained results of intra-view constraint propagation to adjust the weight matrix, which is thus called as constrained weight adjustment (CWA). According to the convergence analysis in Section II-B, we construct a k-NN graph over $\mathcal{X}$ to speed up our intra-view constraint propagation.

1) Intra-View Constraint Propagation: We have just provided a sound solution to the challenging problem of intra-view constraint propagation in Section II. In this subsection, we further consider pairwise constraint propagation over a single view, where each pairwise constraint is defined over a pair of data points from the same view. In fact, this intra-view constraint propagation problem can also be solved from a semi-supervised learning viewpoint by limiting our inter-view constraint propagation to a single view.

Given the dataset $\mathcal{X} = \{x_1, ..., x_N\}$, we denote the set of initial must-link constraints as $\mathcal{M}_\mathcal{X} = \{(x_i, x_j) : l_i = l_j\}$ and the set of initial cannot-link constraints as $\mathcal{C}_\mathcal{X} = \{(x_i, x_j) : l_i \neq l_j\}$, where l_i is the label of data point x_i. Similar to our representation of the initial inter-view pairwise constraints, we first denote the initial intra-view pairwise constraints $\mathcal{M}_\mathcal{X}$ and $\mathcal{C}_\mathcal{X}$ with a single matrix $Z_\mathcal{X} = \{z_{ij}^{(x)}\}_{N \times N}$:

$$z_{ij}^{(x)} = \begin{cases} +1, & (x_i, x_j) \in \mathcal{M}_\mathcal{X}; \\ -1, & (x_i, x_j) \in \mathcal{C}_\mathcal{X}; \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

Furthermore, by making vertical and horizontal observations on $Z_\mathcal{X}$, we further decompose the intra-view constraint propagation problem into semi-supervised learning subproblems, just as our interpretation of inter-view constraint propagation from a semi-supervised learning viewpoint. These subproblems can be similarly merged to a single optimization problem (similar to [20]–[22]):

$$\min_{F_v, F_h} \|F_v - Z_\mathcal{X}\|_{fro}^2 + \mu \text{tr}(F_v^T \mathcal{L}_\mathcal{X} F_v) + \|F_h - Z_\mathcal{X}\|_{fro}^2$$
$$+ \mu \text{tr}(F_h \mathcal{L}_\mathcal{X} F_h^T) + \gamma \|F_v - F_h\|_{fro}^2, \quad (9)$$

where $\mu > 0$ (or $\gamma > 0$) denotes the regularization parameter, and $\mathcal{L}_\mathcal{X}$ denotes the normalized Laplacian matrix defined over the k-NN graph. The second and fourth terms of the above equation denote the energy functional [10] (or the smoothness measure) defined over $\mathcal{X}$. In summary, we have also formulated intra-view constraint propagation as minimizing a regularized energy functional.

Similar to what we have done for solving equation (2), we can adopt the alternate optimization technique to find the best solution to the above intra-view constraint propagation problem. Let $W_\mathcal{X}$ denote the weight matrix of the k-NN graph constructed over the dataset $\mathcal{X}$. The proposed algorithm for our intra-view constraint propagation is outlined as follows:

(1) Compute $S_\mathcal{X} = D^{-\frac{1}{2}} W_\mathcal{X} D^{-\frac{1}{2}}$, where D is a diagonal matrix with its entry (i,i) being the sum of row i of $W_\mathcal{X}$;

(2) Initialize $F_v(0) = 0$, $F_h^* = 0$, and $F_h(0) = 0$;

(3) Iterate $F_v(t+1) = \alpha S_\mathcal{X} F_v(t) + (1-\alpha)((1-\beta)Z_\mathcal{X} + \beta F_h^*)$ until convergence at F_v^*, where $\alpha = \mu/(1 + \mu + \gamma)$ and $\beta = \gamma/(1+\gamma)$;

(4) Iterate $F_h(t+1) = \alpha F_h(t) S_\mathcal{X} + (1-\alpha)((1-\beta)Z_\mathcal{X} + \beta F_v^*)$ until convergence at F_h^*;

(5) Iterate Steps (3)–(4) until the stopping condition is satisfied, and obtain $F^* = (F_v^* + F_h^*)/2$.

(6) Output the normalized solution $F^* = F^*/F_{max}^*$, where F_{max}^* denotes the maximum entry of F^*.

In the experiments, we find that Steps (3)–(5) generally converge in very limited iterations (<10). Moreover, based on k-NN graph, our algorithm has a time cost of $O(kN^2)$ proportional to the number of all possible pairwise constraints. Hence, it can be considered to provide an efficient solution.

2) Weight Adjustment Using Propagated Constraints: It should be noted that the normalized output $F^* = \{f_{ij}^*\}_{N \times N}$ of our intra-view constraint propagation represents an exhaustive set of intra-view pairwise constraints. Our original motivation is to construct a new graph over $\mathcal{X}$ that is fully consistent with F^*. In fact, we can exploit F^* for such graph construction by adjusting the original normalized weight matrix $W_\mathcal{X}$ (i.e. $0 \leq w_{ij}^{(x)} \leq 1$) just as [20]:

$$\tilde{w}_{ij}^{(x)} = \begin{cases} 1 - (1 - f_{ij}^*)(1 - w_{ij}^{(x)}), & f_{ij}^* \geq 0; \\ (1 + f_{ij}^*)w_{ij}^{(x)}, & f_{ij}^* < 0. \end{cases} \quad (10)$$

Since $\tilde{W}_\mathcal{X} = \{\tilde{w}_{ij}^{(x)}\}_{N \times N}$ is nonnegative and symmetric, we then use it as the new weight matrix. Moreover, we can find that $\tilde{W}_{ij}^{(x)} \geq W_{ij}^{(x)}$ (or $< W_{ij}^{(x)}$) if $F_{ij}^* \geq 0$ (or < 0). That is, the new weight matrix $\tilde{W}_\mathcal{X}$ is derived from the original weight matrix $W_\mathcal{X}$ by increasing $W_{ij}^{(x)}$ for the must-link constraints with $F_{ij}^* > 0$ and decreasing $W_{ij}^{(x)}$ for the cannot-link constraints with $F_{ij}^* < 0$. This is entirely consistent with our original motivation of exploiting intra-view pairwise constraints for graph construction.

Once we have constructed the new weight matrix $\tilde{W}_\mathcal{X}$ over $\mathcal{X}$, we can similarly construct the new weight matrix $\tilde{W}_\mathcal{Y}$ over $\mathcal{Y}$. Based on these two new weight matrices, our inter-view constraint propagation can be performed with constrained graph construction (CGC) (as shown in Fig. 1) using constrained weight adjustment (CWA) developed here.

B. Constrained Sparse Representation

The second constrained graph construction method formulates graph construction as sparse representation [23]–[25] and then directly add the intra-view pairwise constraints into sparse representation, which is thus called as constrained sparse representation (CSR). Our work is mainly inspired by recent effort to exploit sparse representation for graph construction, i.e., L_1-graph construction [26], [27]. The basic idea of L_1-graph construction is to seek a sparse linear reconstruction of each data point with the other data points. However, such L_1-graph construction may become infeasible since it incurs

too much time cost given a large data size N. Hence, we only consider the k nearest neighbors of each data point for its sparse linear reconstruction, which thus becomes a much smaller scale optimization problem ($k \ll N$). More notably, due to such neighborhood limitation, the obtained L_1-graph is actually a variant of k-NN graph, which can ensure that our inter-view constraint propagation proposed in Section II runs efficiently on large datasets. Finally, to exploit intra-view pairwise constraints for L_1-graph construction, we seek a constrained sparse linear reconstruction of each data point.

1) L_1-Graph Construction with Sparse Representation: We start with the problem formulation for sparse linear reconstruction of each data point in its k-nearest neighborhood. Given a data point $x_i \in \mathcal{X}$, we suppose it can be reconstructed using its k-nearest neighbors (their indices are collected into $\mathcal{N}_k(i)$), which results in an underdetermined linear system: $x_i = B_i\alpha_i$, where $\alpha_i \in R^k$ is a vector that stores unknown reconstruction coefficients, and $B_i = [x_j]_{j \in \mathcal{N}_k(i)}$ is an overcomplete dictionary with k bases. According to [23], if the solution for x_i is sparse enough, it can be recovered by:

$$\min_{\alpha_i} \ ||\alpha_i||_1, \quad \text{s.t.} \ \ x_i = B_i\alpha_i, \qquad (11)$$

where $||\alpha_i||_1$ is the L_1-norm of α_i. Given the kernel (affinity) matrix $A = \{a_{ij}\}_{N \times N}$ computed over $\mathcal{X}$, we make use of the kernel trick and transform the above problem into:

$$\min_{\alpha_i} \ ||\alpha_i||_1, \quad \text{s.t.} \ \ \hat{x}_i = C_i\alpha_i, \qquad (12)$$

where $\hat{x}_i = [a_{ji}]_{j \in \mathcal{N}_k(i)} \in R^k$, $C_i = [a_{jj'}]_{j,j' \in \mathcal{N}_k(i)} \in R^{k \times k}$. In practice, due to the noise in the data, we can reconstruct $\hat{x}_i$ similar to [24]: $\hat{x}_i = C_i\alpha_i + \zeta_i$, where ζ_i is the noise term. The above L_1-optimization problem can then be redefined by minimizing the L_1-norm of both reconstruction coefficients and reconstruction error:

$$\min_{\alpha_i'} \ ||\alpha_i'||_1, \quad \text{s.t.} \ \ \hat{x}_i = C_i'\alpha_i', \qquad (13)$$

where $C_i' = [C_i, I] \in R^{k \times 2k}$ and $\alpha_i' = [\alpha_i^T, \zeta_i^T]^T$. This convex optimization can be solved by general linear programming and has a globally optimal solution.

After we have obtained the reconstruction coefficients for all the data points by the above sparse linear reconstruction, the weight matrix $W_{\mathcal{X}} = \{w_{ij}^{(x)}\}_{N \times N}$ can be defined by:

$$w_{ij}^{(x)} = \begin{cases} |\alpha_i'(j')|, & j \in \mathcal{N}_k(i), j' = \text{index}(j, \mathcal{N}_k(i)); \\ 0, & \text{otherwise}, \end{cases} \qquad (14)$$

where $\alpha_i'(j')$ denotes the j'-th element of the vector α_i', and $j' = \text{index}(j, \mathcal{N}_k(i))$ means that j is the j'-th element of the set $\mathcal{N}_k(i)$. By setting the weight matrix $W_{\mathcal{X}} = (W_{\mathcal{X}} + W_{\mathcal{X}}^T)/2$, we construct a graph $\mathcal{G}_{\mathcal{X}} = \{\mathcal{X}, W_{\mathcal{X}}\}$ over $\mathcal{X}$, which is called as L_1-graph since it is constructed by L_1-optimization.

2) L_1-Norm Laplacian Regularization with Intra-View Pairwise Constraints: In the above L_1-graph construction, we have ignored intra-view pairwise constraints (see examples in Fig. 1). In fact, this supervisory information can be exploited for L_1-graph construction through Laplacian regularization [9], [10]. Our basic idea is to first derive Laplacian regularization from intra-view pairwise constraints and then incorporate

this constrained term into sparse linear reconstruction (the key step of L_1-graph construction). In the following, we will first elaborate how to derive a new Laplacian regularization term from intra-view pairwise constraints.

Given a set of intra-view must-link constraints $\mathcal{M}_{\mathcal{X}}$ and a set of intra-view cannot-link constraints $\mathcal{C}_{\mathcal{X}}$ defined over $\mathcal{X}$, we can represent both $\mathcal{M}_{\mathcal{X}}$ and $\mathcal{C}_{\mathcal{X}}$ using a single matrix $Z_{\mathcal{X}} = \{z_{ij}^{(x)}\}_{N \times N}$ exactly the same as equation (8). The normalized Laplacian matrix limited to the k-nearest neighborhood of data point x_i can thus be defined as:

$$\mathcal{L}_i = I - D_i^{-1/2}(1 + Z_i)D_i^{-1/2}, \qquad (15)$$

where $Z_i = [z_{jj'}^{(x)}]_{j,j' \in \mathcal{N}_k(i)} \in R^{k \times k}$, and D_i is a diagonal matrix with its j-th diagonal element being the sum of the j-th row of $1 + Z_i$. Here, we define the similarity matrix (i.e. $1 + Z_i$) limited to the k-nearest neighborhood $\mathcal{N}_k(i)$ of x_i based on the intra-view pairwise constraints stored in $Z_{\mathcal{X}}$. From this normalized Laplacian matrix $\mathcal{L}_i$, we can derive the Laplacian regularization term for the sparse representation problem (12) as $\alpha_i^T \mathcal{L}_i \alpha_i$, the same as the original definition in [9].

However, we have difficulty in directly incorporating this Laplacian regularization term into the sparse representation problem (12), no matter as a part of the objective function or a constraint condition. Hence, we further formulate an L_1-norm version of Laplacian regularization [12], [28]–[30]:

$$||\tilde{C}_i\alpha_i||_1 = ||\Sigma_i^{\frac{1}{2}}V_i^T\alpha_i||_1, \qquad (16)$$

where $\tilde{C}_i = \Sigma_i^{\frac{1}{2}}V_i^T$, V_i is a $k \times k$ orthonormal matrix with each column being an eigenvector of $\mathcal{L}_i$, and Σ_i is a $k \times k$ diagonal matrix with its diagonal element $\Sigma_i(j,j)$ being an eigenvalue of $\mathcal{L}_i$ (sorted as $\Sigma_i(1,1) \leq ... \leq \Sigma_i(k,k)$). Given that $\mathcal{L}_i$ is nonnegative definite, $\Sigma_i \geq 0$ (i.e. all the eigenvalues ≥ 0). Since $\mathcal{L}_iV_i = V_i\Sigma_i$ and V_i is orthonormal, we have $\mathcal{L}_i = V_i\Sigma_iV_i^T$. Hence, the original Laplacian regularization $\alpha_i^T\mathcal{L}_i\alpha_i$ can be reformulated as:

$$\alpha_i^T\mathcal{L}_i\alpha_i = \alpha_i^TV_i\Sigma_i^{\frac{1}{2}}\Sigma_i^{\frac{1}{2}}V_i^T\alpha_i = ||\tilde{C}_i\alpha_i||_2^2, \qquad (17)$$

which means that our new formulation $||\tilde{C}_i\alpha_i||_1$ can indeed be regarded as an L_1-norm version of the original Laplacian regularization $\alpha_i^T\mathcal{L}_i\alpha_i = ||\tilde{C}_i\alpha_i||_2^2$.

3) L_1-Graph Construction with L_1-Norm Laplacian Regularization: After we have formulated L_1-norm Laplacian regularization based on intra-view pairwise constraints, we can further incorporate this constrained term into sparse linear reconstruction used for L_1-graph construction. More concretely, by introducing noise terms for linear reconstruction and L_1-norm Laplacian regularization, we transform the sparse representation problem (12) into

$$\min_{\alpha_i, \zeta_i, \xi_i} \ ||[\alpha_i^T, \zeta_i^T, \xi_i^T]||_1,$$
$$\text{s.t.} \ \ \hat{x}_i = C_i\alpha_i + \zeta_i, \ 0 = \tilde{C}_i\alpha_i + \xi_i, \qquad (18)$$

where the reconstruction error and Laplacian regularization with respect to α_i are controlled by ζ_i and ξ_i, respectively. Let $\alpha_i' = [\alpha_i^T, \zeta_i^T, \xi_i^T]^T$, $C_i' = \begin{bmatrix} C_i & I & 0 \\ \tilde{C}_i & 0 & I \end{bmatrix}$, and $\hat{x}_i' =$

$[\hat{x}_i^T, 0^T]^T$. We finally solve the following constrained spare representation problem for L_1-graph construction:

$$\min_{\alpha_i'} \quad ||\alpha_i'||_1, \quad \text{s.t.} \quad \hat{x}_i' = C_i' \alpha_i', \qquad (19)$$

which takes the same form as the original spare representation problem (13). Here, it is noteworthy that this constrained spare representation (CSR) problem can be solved very efficiently, since it is limited to k-nearest neighborhood. The weight matrix $W_{\mathcal{X}}$ of the L_1-graph $\mathcal{G}_{\mathcal{X}} = \{\mathcal{X}, W_{\mathcal{X}}\}$ can be defined the same as equation (14).

In our CSR formulation, the L_1-norm Laplacian regularization can be smoothly incorporated into the original sparse representation problem (12). However, this is not true for the traditional Laplacian regularization [9], [10], which may introduce extra parameters (hard to tune in practice) into the L_1-optimization for sparse representation. Meanwhile, our L_1-norm Laplacian regularization can induce another type of sparsity (see the extra noise term ξ_i), which can not be ensured by the traditional Laplacian regularization. Moreover, the p-Laplacian regularization [31] can also be regarded as an ordinary L_1-generalization of the Laplacian regularization when $p = 1$. According to [32], by defining a matrix $C_p \in R^{\frac{k(k-1)}{2} \times k}$, the p-Laplacian regularization can be formulated as $||C_p \alpha_i||_1$, similar to our L_1-norm Laplacian regularization. Hence, we can similarly apply the p-Laplacian regularization with $p = 1$ to constrained spare representation. However, such Laplacian regularization incurs large time cost due to the large matrix C_p even for small neighborhood size (e.g. $k = 90$).

Once we have constructed the L_1-graph $\mathcal{G}_{\mathcal{X}} = \{\mathcal{X}, W_{\mathcal{X}}\}$ over $\mathcal{X}$, we can similarly construct the L_1-graph $\mathcal{G}_{\mathcal{Y}} = \{\mathcal{Y}, W_{\mathcal{Y}}\}$ over $\mathcal{Y}$. Based on the two weight matrices, our inter-view constraint propagation can be performed with constrained graph construction (CGC) (as shown in Fig. 1) using constrained sparse representation (CSR) developed here.

IV. APPLICATION TO CROSS-VIEW RETRIEVAL

When multiple views refer to text, image, audio and so on (see Fig. 3), the output of our inter-view constraint propagation actually can be viewed as the correlation between different media views. As we have mentioned, given the output $F^* = \{f_{ij}^*\}_{N \times M}$ of our inter-view constraint propagation, (x_i, y_j) denotes a must-link (or cannot-link) constraint if $f_{ij}^* > 0$ (or < 0). Considering the inherent meanings of must-link and cannot-link constraints, we can state that: x_i and y_j are "positively correlated" if $f_{ij}^* > 0$, while they are "negatively correlated" if $f_{ij}^* < 0$. Hence, we can view f_{ij}^* as the correlation coefficient between x_i and y_j. The distinct advantage of such interpretation of F^* as a correlation measure is that F^* can thus be used for ranking on $\mathcal{Y}$ given a query x_i or ranking on $\mathcal{X}$ given a query y_j. In fact, this is just the goal of cross-view retrieval which has drawn much attention recently [15]. That is, such task can be directly handled by our inter-view constraint propagation.

In this paper, we focus on a special case of cross-view retrieval, i.e. only text and image views are considered. In this case, cross-view retrieval is somewhat similar to automatic image annotation [33]–[36] and image caption generation
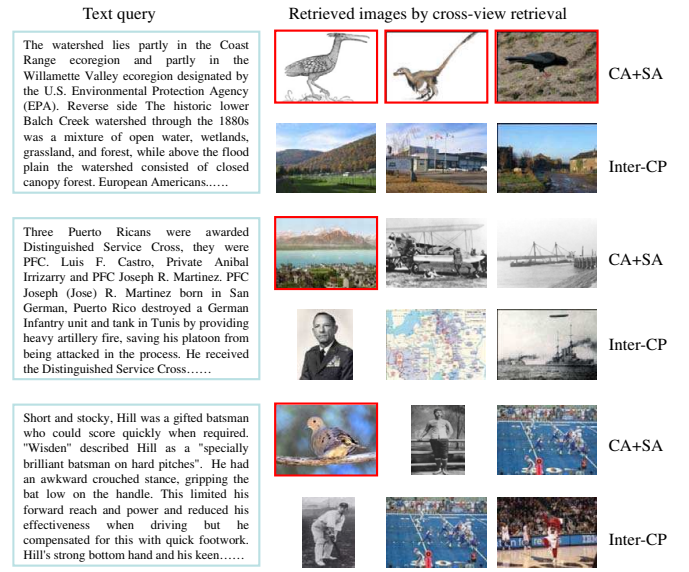

Fig. 3. Cross-view retrieval examples on the Wikipedia benchmark dataset [15]. Here, the incorrectly retrieved images are marked with red boxes.

[37]–[39], since these three tasks all aim to learn the relations between the text and image views. However, even if only text and image views are considered, cross-view retrieval is still quite different from automatic image annotation and image caption generation. More concretely, automatic image annotation relies on very limited types of textual representations and mainly associates images only with textual keywords, while cross-view retrieval is designed to deal with much more richly annotated data, motivated by the ongoing explosion of Web-based content such as news archives and Wikipedia pages. Similar to cross-view retrieval, image caption generation can also deal with more richly annotated data (i.e. captions) with respect to the textual keywords concerned in automatic image annotation. However, this task tends to model image captions as sentences by exploiting certain prior knowledge (e.g. the <object, action, scene> triplets used in [37]), different from cross-view retrieval that focuses on associating images with complete text articles using no prior knowledge from the text view (any general textual representations are applicable actually once their similarities are provided).

In the context of cross-view retrieval, one notable recent work is [15] which first learns the correlation between the text and image views with canonical correlation analysis (CCA) [40] and then achieves the abstraction by representing text and image at a more general semantic level. However, two separate steps, i.e. correlation analysis (CA) and semantic abstraction (SA), are involved in this modeling, and the use of semantic abstraction after CCA (i.e. CA+SA) seems rather ad hoc. Fortunately, this problem can be completely addressed by our inter-view constraint propagation (Inter-CP). The semantic information (e.g. class labels) associated with images and text can be used to define the initial must-link and cannot-link constraints based on the training dataset, while the correlation between text and image views can be explicitly learnt by the proposed algorithm in Section II. That is, the correlation analysis and semantic abstraction has been successfully in-
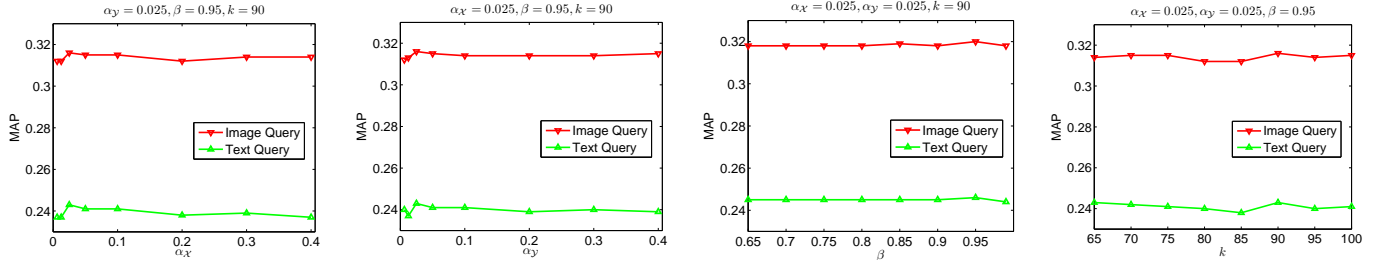
Fig. 4. The cross-view retrieval results by cross-validation on the training set of the Wikipedia dataset for our Inter-CP algorithm (CSR is used here).

tegrated in our inter-view constraint propagation framework. The effectiveness of such integration as compared to CA+SA [15] is preliminarily verified by several cross-view retrieval examples shown in Fig. 3. Further verification will be provided in our later experiments. More notably, although only tested in cross-view retrieval, our inter-view constraint propagation can be readily extended to other multi-view tasks, since it has actually learnt the correlation between different views.

V. EXPERIMENTAL RESULTS

In this section, our inter-view constraint propagation (Inter-CP) algorithm is evaluated in the challenging application of cross-view retrieval. We focus on comparing our Inter-CP algorithm with the state-of-the-art approach [15], since they both consider not only correlation analysis (CA) but also semantic abstraction (SA) for text and image views. Moreover, we also make comparison with another two closely related approaches that integrate CA and SA for cross-view retrieval similar to [15] but perform correlation analysis by partial least squares (PLS) [41] and cross-modal factor analysis (CFA) [42] instead of CCA, respectively. In the following, these two CA+SA approaches are denoted as CA+SA (PLS) and CA+SA (CFA), while the state-of-the-art approach [15] is denoted as CA+SA (CCA). Finally, to show the effectiveness of constrained graph construction, we construct four types of graphs for our Inter-CP algorithm: k-NN graph (k-NN), L_1-graph using sparse representation (SR), k-NN graph using constrained weight adjustment (CWA), and L_1-graph using constrained sparse representation (CSR).

A. Experimental Setup

We select two different datasets for performance evaluation. The first one is a Wikipedia benchmark dataset [15], which contains a total of 2,866 documents derived from Wikipedia's "featured articles". Each document is actually a text-image pair, annotated with a label from the vocabulary of 10 semantic classes. This benchmark dataset [15] is split into a training set of 2,173 documents and a test set of 693 documents. Moreover, the second dataset consists of totally 8,564 documents crawled from the photo sharing website Flickr. The image and text views of each document denote a photo and a set of tags provided by the users, respectively. Although such text presentation does not take a free form as that for the Wikipedia dataset, it is rather noisy since many of the tags may be incorrectly annotated by the users. This Flickr dataset is organized into 11 semantic classes. We split it into a training set of 4,282 documents and a test set of the same size.

For the above two datasets, we take the same strategy as [15] to generate both text and image representation. More concretely, in the Wikipedia dataset, the text representation for each document is derived from a latent Dirichlet allocation model with 10 latent topics, while the image representation is based on a bag-of-words model with 128 visual words learnt from the extracted SIFT descriptors, just as [15]. Moreover, for the Flickr dataset, we generate the text and image representation similarly, and the main difference is that we select a relatively large visual vocabulary (of the size 2,000) for image representation and refine the noisy textual vocabulary to the size 1,000 by a preprocessing step for text representation.

In our experiments, the intra-view pairwise constraints used for our CGC and inter-view pairwise constraints used for our Inter-CP are initially derived from the class labels of the training documents of each dataset. The performance of our Inter-CP with CGC is evaluated on the test set. Here, two tasks of cross-view retrieval are considered: text retrieval using an image query, and image retrieval using a text query. In the following, these two tasks are denoted as "Image Query" and "Text Query", respectively. For each task, the retrieval results are measured with mean average precision (MAP) which has been widely used in the image retrieval literature [18].

Let $\mathcal{X}$ denote the text representation and $\mathcal{Y}$ denote the image representation. For our Inter-CP algorithm, we perform CGC over $\mathcal{X}$ and $\mathcal{Y}$ with the same k. The parameters of our Inter-CP algorithm with CGC can be selected by fivefold cross-validation on the training set. For example, according to Fig. 4, we set the parameters of our Inter-CP (CSR is used for CGC) on the Wikipedia dataset as: $\alpha_{\mathcal{X}} = 0.025$, $\alpha_{\mathcal{Y}} = 0.025$, $\beta = 0.95$, and $k = 90$. It is noteworthy that our Inter-CP with CSR is not sensitive to these parameters. Moreover, the parameters of our Inter-CP with CWA can be similarly set to their respective optimal values. To summarize, we have selected the best values for all the parameters of our UCP algorithm with CGC by cross-validation on the training set. For fair comparison, we take the same parameter selection strategy for other closely related algorithms.

B. Retrieval Results

The cross-view retrieval results on the two datasets are listed in Tables I and II, respectively. The immediate observation is that we can achieve the best results when both intra-view and inter-view pairwise constraints are exploited by Inter-CP+CWA (or Inter-CP+CSR). This means that our Inter-CP with CGC can most effectively exploit the initial supervisory information provided for cross-view retrieval. As compared

TABLE I
THE CROSS-VIEW RETRIEVAL RESULTS ON THE TEST SET OF THE WIKIPEDIA DATASET MEASURED BY THE MAP SCORES.

Methods	Image Query	Text Query	Average
CA+SA (PLS)	0.250	0.190	0.220
CA+SA (CFA)	0.272	0.221	0.247
CA+SA (CCA)	0.277	0.226	0.252
Inter-CP+k-NN	0.329	0.256	0.293
Inter-CP+SR	0.336	0.259	0.298
Inter-CP+CWA	0.337	0.260	0.299
Inter-CP+CSR	**0.343**	**0.268**	**0.306**

TABLE II
THE CROSS-VIEW RETRIEVAL RESULTS ON THE TEST SET OF THE FLICKR DATASET MEASURED BY THE MAP SCORES.

Methods	Image Query	Text Query	Average
CA+SA (PLS)	0.201	0.168	0.185
CA+SA (CFA)	0.252	0.231	0.242
CA+SA (CCA)	0.280	0.263	0.272
Inter-CP+k-NN	0.495	0.483	0.489
Inter-CP+SR	0.509	0.496	0.503
Inter-CP+CWA	0.521	0.499	0.510
Inter-CP+CSR	**0.521**	**0.505**	**0.513**

to the three CA+SA approaches by semantic abstraction after correlation analysis (via PLS, CFA, or CCA), our Inter-CP can seamlessly integrate these two separate steps and then lead to much better results. Moreover, the effectiveness of our CGC is verified by the comparison Inter-CP+CWA vs. Inter-CP+k-NN (or Inter-CP+CSR vs. Inter-CP+SR), especially on the Flickr dataset. As for our two CGC methods, CSR is shown to perform better than CWA, which is mainly due to the noise-robustness property of sparse representation.

It should be noted that our Inter-CP algorithm can be considered to provide an efficient solution, since it has a time cost proportional to the number of all possible pairwise constraints. This is also verified by our observations in the experiments. For example, the running time taken by CA+SA (CCA, CFA or PLS), Inter-CP+k-NN, and Inter-CP+CWA on the Wikipedia dataset is 10, 24, and 55 seconds, respectively. Here, we run all the algorithms (Matlab code) on a computer with 3GHz CPU and 32GB RAM. Since our Inter-CP with CGC leads to significantly better results, we prefer it to CA+SA in practice, regardless of its relatively larger time cost.

VI. Conclusions

In this paper, we have investigated the challenging problem of pairwise constraint propagation on multi-view data. By decomposing the inter-view constraint propagation problem into a set of independent semi-supervised learning subproblems, we have uniformly formulated them as minimizing a regularized energy functional. More importantly, these semi-supervised learning subproblems can be solved efficiently using label propagation with k-NN graph. We then develop two constrained graph construction methods for our inter-view constraint propagation, and the obtained two graphs can be considered as the variants of k-NN graph. The experimental results in cross-view retrieval have shown the promising performance of our inter-view constraint propagation with constrained graph construction. For future work, our method will be extended to other multi-view tasks.

References

[1] Z. Lu and H. H.-S. Ip, "Generalized competitive learning of gaussian mixture models," *IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 39, no. 4, pp. 901–909, 2009.

[2] Z. Lu, "An iterative algorithm for entropy regularized likelihood learning on gaussian mixture with automatic model selection," *Neurocomputing*, vol. 69, no. 13, pp. 1674–1677, 2006.

[3] L. Wang, Z. Lu, and H. Ip, "Image categorization based on a hierarchical spatial markov model," in *CAIP*, 2009, pp. 766–773.

[4] Z. Lu, Y. Peng, and J. Xiao, "From comparing clusterings to combining clusterings," in *AAAI*, 2008, pp. 665–670.

[5] Z. Lu and M. Carreira-Perpinan, "Constrained spectral clustering through affinity propagation," in *CVPR*, 2008, pp. 1–8.

[6] Z. Li, J. Liu, and X. Tang, "Pairwise constraint propagation by semidefinite programming for semi-supervised classification," in *ICML*, 2008, pp. 576–583.

[7] S. Yu and J. Shi, "Segmentation given partial grouping constraints," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 2, pp. 173–183, 2004.

[8] S. Kamvar, D. Klein, and C. Manning, "Spectral learning," in *IJCAI*, 2003, pp. 561–566.

[9] D. Zhou, O. Bousquet, T. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," in *Advances in Neural Information Processing Systems 16*, 2004, pp. 321–328.

[10] X. Zhu, Z. Ghahramani, and J. Lafferty, "Semi-supervised learning using Gaussian fields and harmonic functions," in *ICML*, 2003, pp. 912–919.

[11] Z. Lu and H. Ip, "Image categorization by learning with context and consistency," in *CVPR*, 2009, pp. 2719–2726.

[12] Z. Lu and L. Wang, "Noise-robust semi-supervised learning via fast sparse coding," *Pattern Recognition*, vol. 48, no. 2, pp. 605–612, 2015.

[13] Z. Lu and Y. Peng, "Combining latent semantic learning and reduced hypergraph learning for semi-supervised image categorization," in *ACM Multimedia*, 2011, pp. 1409–1412.

[14] Z. Lu and H. Ip, "Combining context, consistency, and diversity cues for interactive image categorization," *IEEE Trans. Multimedia*, vol. 12, no. 3, pp. 194–203, 2010.

[15] N. Rasiwasia, J. Costa Pereira, E. Coviello, G. Doyle, G. Lanckriet, R. Levy, and N. Vasconcelos, "A new approach to cross-modal multimedia retrieval," in *ACM Multimedia*, 2010, pp. 251–260.

[16] E. Eaton, M. desJardins, and S. Jacob, "Multi-view clustering with constraint propagation for learning with an incomplete mapping between views," in *CIKM*, 2010, pp. 389–398.

[17] Z. Fu, H. Ip, H. Lu, and Z. Lu, "Multi-modal constraint propagation for heterogeneous image clustering," in *ACM Multimedia*, 2011, pp. 143–152.

[18] M. Guillaumin, J. Verbeek, and C. Schmid, "Multimodal semi-supervised learning for image classification," in *CVPR*, 2010, pp. 902–909.

[19] E. Bruno, N. Moenne-Loccoz, and S. Marchand-Maillet, "Design of multimodal dissimilarity spaces for retrieval of video documents," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 30, no. 9, pp. 1520–1533, 2008.

[20] Z. Lu and H. Ip, "Constrained spectral clustering via exhaustive and efficient constraint propagation," in *ECCV*, vol. 6, 2010, pp. 1–14.

[21] Z. Fu, Z. Lu, H. H.-S. Ip, Y. Peng, and H. Lu, "Symmetric graph regularized constraint propagation," in *AAAI*, 2011.

[22] Z. Lu and Y. Peng, "Exhaustive and efficient constraint propagation: A graph-based learning approach and its applications," *International Journal of Computer Vision*, vol. 103, no. 3, pp. 306–325, 2013.

[23] D. Donoho, "For most large underdetermined systems of linear equations the minimal ℓ^1-norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics*, vol. 59, no. 7, pp. 797–829, 2004.

[24] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[25] Z. Lu and Y. Peng, "Latent semantic learning by efficient sparse coding with hypergraph regularization." in *AAAI*, 2011.

[26] H. Cheng, Z. Liu, and J. Yang, "Sparsity induced similarity measure for label propagation," in *ICCV*, 2009, pp. 317–324.

[27] B. Cheng, J. Yang, S. Yan, and T. Huang, "Learning with ℓ^1-graph for image analysis," *IEEE Trans. Image Processing*, vol. 19, no. 4, pp. 858–866, Apr. 2010.

[28] Z. Lu and Y. Peng, "Image annotation by semantic sparse recoding of visual content," in *ACM Multimedia*, 2012, pp. 499–508.

[29] ——, "Latent semantic learning with structured sparse representation for human action recognition," *Pattern Recognition*, vol. 46, no. 7, pp. 1799–1809, 2013.

[30] Z. Lu, P. Han, L. Wang, and J.-R. Wen, "Semantic sparse recoding of visual content for image applications," *IEEE Trans. Image Processing*, vol. 24, no. 1, 2015.

[31] D. Zhou and B. Schölkopf, "Regularization on discrete spaces," in *DAGM*, 2005, pp. 361–368.

[32] X. Chen, Q. Lin, S. Kim, J. G. Carbonell, and E. P. Xing, "Smoothing proximal gradient method for general structured sparse learning," in *UAI*, 2011, pp. 105–114.

[33] J. Li and J. Wang, "Automatic linguistic indexing of pictures by a statistical modeling approach," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1075–1088, Sept. 2003.

[34] S. Feng, R. Manmatha, and V. Lavrenko, "Multiple Bernoulli relevance models for image and video annotation," in *CVPR*, vol. 2, 2004, pp. 1002–1009.

[35] Z. Lu, H. Ip, and Y. Peng, "Contextual kernel and spectral methods for learning the semantics of images," *IEEE Trans. Image Processing*, vol. 20, no. 6, pp. 1739–1750, 2011.

[36] Z. Lu and H. Ip, "Spatial markov kernels for image categorization and annotation," *IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 41, no. 4, pp. 976–989, 2011.

[37] A. Farhadi, M. Hejrati, M. Sadeghi, P. Young, C. Rashtchian, J. Hockenmaier, and D. Forsyth, "Every picture tells a story: Generating sentences from images," in *ECCV*, vol. 4, 2010, pp. 15–29.

[38] G. Kulkarni, V. Premraj, S. Dhar, S. Li, Y. Choi, A. Berg, and T. Berg, "Baby talk: Understanding and generating simple image descriptions," in *CVPR*, 2011, pp. 1601–1608.

[39] V. Ordonez, G. Kulkarni, and T. Berg, "Im2Text: Describing images using 1 million captioned photographs," in *Advances in Neural Information Processing Systems 24*, 2012, pp. 1143–1151.

[40] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3-4, pp. 321–377, 1936.

[41] H. Wold, "Partial least squares," in *Encyclopedia of Statistical Sciences*, S. Kotz and N. Johnson, Eds. New York: Wiley, 1985, pp. 581–591.

[42] D. Li, N. Dimitrova, M. Li, and I. K. Sethi, "Multimedia content processing through cross-modal association," in *ACM Multimedia*, 2003, pp. 604–611.